DOCUMENT RESUME

ED 336 722 CS 010 700

AUTHOR Williams, Trevor
TITLE IEA Reading Literacy Study--Policy Issues: National and International Perspectives.
PUB DATE Apr 91
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (72nd, Chicago, IL, April 3-7, 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Research; Elementary Secondary Education; Higher Education; *Literacy; *National Surveys; *Reading Achievement; *Research Design; *Research Methodology; Student Evaluation
IDENTIFIERS *International Assn Evaluation Educ Achievement; *United States

ABSTRACT

Mainly describing research methodology, this paper outlines an analysis and reporting strategy for the United States national analyses of reading literacy. The first section of the paper is a brief description of the IEA Reading Literacy study and the issues it seeks to answer, along with an outline of three constraints which condition the development of an analysis and reporting strategy. The second section raises issues about the specification of the audience for the findings of the study, and about the nature of the reporting most suitable for this audience. The third section outlines the kind of theoretical/substantive decisions that must be made before analyses can begin, decisions about constructs and models, and presents the IEA Reading Literacy model. The fourth section provides a discussion concerning the presentation of descriptive statistics from the study, and proposes a methodology for reporting multivariate statistics in a form acceptable to a general audience. The fifth section outlines a reporting strategy designed to make the findings of the study accessible to a general audience. The last section provides an outline of the structure and substance of the first United States national report. (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED336722

IEA READING LITERACY STUDY

Policy Issues: National and International Perspectives

AERA Annual Meetings
April 1991

Trevor Williams
Australian Council for Educational Research

DRAFT: please do not cite or quote

CS010700

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Trevor Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

A. Introduction

This is mainly a methodological paper. Its main task is to outline an analysis and reporting strategy for the U.S. national analyses. However, substantive matters in the study of reading are taken up in the final section where we outline the structure and content proposed for the U.S. national report.

The paper is divided into six sections. Section A, the present section, provides a brief description of the IEA Reading Literacy study and the issues it seeks to answer, along with an outline of three constraints which condition the development of an analysis and reporting strategy. Section B raises issues about the specification of the audience for the findings of this study, and about the nature of the reporting most suitable for this audience. Section C outlines the kind of theoretical/substantive decisions that must be made before analyses can begin, decisions about constructs and models, and presents our view of the IEA Reading Literacy model as it now stands. Section D provides a discussion concerning the presentation of descriptive statistics from the study, and proposes a methodology for reporting multivariate statistics in a form acceptable to a general audience. Section E outlines a reporting strategy designed to make the findings of the study accessible--and hence, useful--to a general audience. Section F provides and outline of our thinking about the structure and substance of the first U.S. national report.

The proposals advanced here probably have application to the analyses planned by other nations, and to the international analyses, but no attempt is made to develop the argument in these directions except to note the more obvious implications.

1. The IEA Reading Literacy Study.

This section provides a brief overview of the project, concentrating on the basic issues driving the international study, and those which are to shape the U.S. analyses, along with the parameters of the samples which will provide the data for these analyses. And, as a foundation for the discussion which follows, consideration is given also to the constraints within which these analyses must be designed and executed.

1.1. The IEA International Analyses.

The International analyses center around:

a. between-nation comparisons of reading achievement levels;

b. between-nation comparisons of influences on reading achievement--essentially comparisons of within-nation analyses such as those described below; and

c. the relationship between national economic indicators and national reading literacy levels.

1.2. The Focus of the U.S. Analyses.

The main thrust of the U.S. analyses is to be the explanation of individual differences in each of the three dimensions of reading literacy--narrative,

expository and document. Several sources of these differences are considered though not all are seen to be of equal importance for the present analyses. The issues to be addressed, in order of importance, are as follows:

a. effects of school instructional strategies;

b. effects of school, teacher and program characteristics;

c. effects of family instructional variables;

d. effects of ascribed family characteristics (SES, ethnicity, etc.); and

e. effects of family-school interaction;

These effects on reading literacy are seen to be mediated by student reading strategies and by student reading activities. In other words, the effects of school instructional strategies on reading achievement come about, in part, because these instructional strategies affect the reading strategies and activities of students which, in turn, influence reading achievement.

1.3. The U.S. Sample.

National probability samples of fourth grade and nineth grade students will provide the data on which the analyses are to be based.

2. Design Constraints.

In common with most IEA studies, the Reading Literacy study does not have a tight design. This fact imposes three important constraints on the design of an analysis and reporting strategy for the U.S. study. We mention them here because they condition much of the following discussion about strategy.

a. The study was designed within a fairly general theoretical framework.

b. For at least this reason, questions developed for inclusion in the study (by committees of NRCs) tend not to have explicit connections with explicit constructs since only rather general construct categories were defined in advance of the item definitions. (The reading achievement tests are something of an exception.)

c. The basic parameters of the data are fixed by international concensus, although national options have been included to the extent allowed by the time available for data collection.

This kind of design may be the price one pays to gain the cooperation and to coordinate the activities of 30 or 40 nations in such an enterprise. Whatever, the analysis plan must operate within this context, and this fact has important implications for how one approaches the design. The overriding constraint is that the variables and their measures are fixed but are only loosely grouped into 'theoretical' categories defined somewhat imprecisely. Decisions about model and construct development at the operational level must be made within the boundaries of this defined set of

variables, and must take into account the established measurement properties of these variables. In many ways the design issues are those faced in traditional secondary analyses of survey data.

B. POLICY RESEARCH FOR A POLICY AUDIENCE

Before the design of the analysis strategy can proceed the primary audience for the reports needs to be identified. With the readers clearly defined the forms of analysis and reporting can be tailored appropriately.

1. The IEA Reading Literacy Study as 'Policy Research'.

We assume at the outset that the IEA studies as a whole are probably developed more with a view to providing information to influence educational policy than as a means of advancing discipline theory/knowledge. In other words, the present IEA Reading Literacy study has been shaped as 'policy research' rather than 'discipline research'. While the data may well provide for analyses that contribute to discipline theory and knowledge, these will be a secondary consideration in the design of the analysis and reporting strategy. Of primary importance is the reporting of the results in such a way that they will be available to those involved with the formulation of educational policy.

1.1. Discipline and Policy Research.

This assumption shapes the overall strategy presented in this paper. At a fundamental level the strategy is influenced by James Coleman's notions about the distinction between discipline and policy research. This distinction hinges around a difference in purpose. *Discipline research* is initiated and designed within the discipline with the view to contributing to discipline theory and knowledge. It is carried out using the methods and statistics of the discipline, and reported in these terms on the assumption that the audience--members of the discipline--will be familiar with the concepts and techniques. *Policy research* is initiated in the 'world of action', as Coleman puts it, outside the discipline. Further, its findings are destined for the world of action as well, as information which will contribute to the formation of social policy. However, the research itself is carried out within the discipline environment using the theoretical frameworks, methods and statistics of the discipline.

1.2. The Translation Issue.

It follows that a critical aspect of the design and execution of policy research is the matter of translation: of the problem, as defined in the world of action, into the conceptual schemes of the discipline so that the research can be undertaken; and, of the findings, as defined in the discipline, into a form suitable for assimilation by a general audience most of whom will be unfamiliar with the language and techniques of the discipline.

The analysis and reporting strategy suggested here does not dwell at length on the question of how the discipline research component of the Reading Literacy study should be undertaken. Rather, we focus on the twin translation issues. Consideration of the question of 'translation in' leads

us to ask questions about the nature of the specific constructs proposed, the measurement models which link indicators to these constructs, the nature of the proposed relationships among constructs, and how the questions which drive the study will be answered by these formulations.

'Translation out' issues, those to do with the translation of the findings out of the discipline into a form suitable for a general policy audience, are covered in proposals for particular forms of statistical analysis and reporting.

2. The Audience

Since the Reading Literacy study is being seen as policy research we have defined the primary audience as a general audience--practitioners, policy analysts, politicians, academics, and the general (interested) public. However, even with such a non-specialist audience as the focus for the reporting the obligation to expose in full detail the basis of the findings reported remains. We outline below some strategies for achieving this end.

2.1. Reporting Statistics for a General Audience

The first problem to be faced has to do with the reporting of statistics--descriptive and multivariate. Each type of statistic has its own problems and advantages. The statistical displays in descriptive analyses are readily understood, but the substantive inferences that one can make are limited by an inability to control on more than one or two variables. In the case of multivariate analyses, the substantive interpretations are less equivocal because more extraneous influences can be held constant, but the statistics produced are usually outside the experience of a general audience.

Reports written for general audiences commonly provide only descriptive statistics for the reasons noted above. These reports suffer as a result because they provide equivocal interpretations of the data (or none at all), and invite misinterpretations on the part of the reader. At the other extreme, 'academic' presentations offer less equivocal interpretations on the basis of their multivariate analyses but tend toward the impenetrable from the point of view of a lay audience.

If we assume that reasonably unequivocal interpretations of the data are required then, in addition to the reporting of fairly simple descriptive statistics, multivariate analyses must be undertaken and their findings reported. The question is one of how to do this.

2.2.1. General and Technical Reports.

One approach would be to produce separate 'general' and 'technical' reports, the former containing descriptive statistical presentations accompanied by interpretations based on the multivariate analyses in the technical report. The technical reports could adopt a standard academic format. The problem with this strategy is that interpretations (actual or implied) of the descriptive data presented in the general report will often not match the conclusions reached from the multivariate analyses, as one would expect. Moreover, the conclusions presented have to be taken on faith by members of a general audience since the analyses on which they are based are presented elsewhere and in a form not readily comprehended.

2.2.2. A 'Technical' Report for a General Audience.

Since there is a little of the methodologist in most people, a more desirable approach is to make each report reasonably self-contained by reporting the results of both descriptive and multivariate analyses, and the interpretations that follow. However, this will not go a long way toward satisfying the needs of the audience unless the multivariate statistics themselves are made comprehensible. Appending the technical report to the report for the general audience is not a totally satisfactory solution. What is needed is a strategy for presenting the results of multivariate analyses in a form that the members of a general audience will accept and understand.

We suggest a strategy in section D below, based on the following assumptions:

a. a general audience will not be familiar with the literature or technical terms associated with the study of reading, but these are not of such an esoteric nature that they cannot be explained in lay terms;

b. multivariate statistics are necessary to provide reasonably unequivocal interpretations about family, school and instructional influences on reading literacy;

c. the principal form of multivariate analysis used in estimating the IEA Reading Literacy models will be a regression-like technique producing something like partial slope coefficients to be interpreted as measures of effect;

d. a general audience will not be familiar with multivariate statistics of this kind as they are usually presented and, no matter how well explained, these summary statistics have little intuitive meaning for such an audience;

e. a general audience will be familiar with the statistics of common discourse, means and percentages;

f. a report which offers integrated empirical support, in the form of data displays, for the conclusions advanced in its text is likely to have a greater impact on its audience than one which relies on a text alone.

2.2.3. Technical Reports.

It is probably sensible to think of several purely technical reports as well to cater for those who will want to explore the findings in detail, and to serve as general technical references for the study: a sampling design report, one or more volumes reporting the multivariate analyses in their complete technical detail, a compendium of questionnaires and tests, a report detailing the item analyses for the cognitive measures, the usual 'telephone book' of response distributions item by item, and so on.

C. DEFINING MODELS AND CONSTRUCTS

This section addresses 'theoretical' issues to do with the translation of the central research questions which focus the Reading Literacy study,

questions defined essentially outside the discipline, into forms consistent discipline research in the area of reading. Attention is directed to the need to develop models of constructs, and of the relations between constructs, that are consistent with the theory and substance of the study of reading. Attention is also directed to analysis issues which arise from the need to collect data and provide statistics which address the basic questions underlying the study.

1. Theories, Causes, Statistics and Effects.

The discussion which follows takes the position that causal thinking about the relations between constructs and their indicators, and about the relations between the constructs themselves, is the only sensible way to proceed in defining, and talking about, workable constructs and models to guide the analyses. People think about reading literacy, especially about those things that make a difference to children's competencies in this respect, in causal terms. Theoretical thinking is causal thinking. The models discussed below are presented in these terms.

In short, the statistics presented **must** lend themselves to a cause-effect interpretation or they are of little interest to anybody. There is no way around this situation. As soon as one moves from univariate statistics to bivariate statistics and beyond, causal interpretations come into play. (Straightout head-counting and atheoretical prediction exercises are the exceptions.) They may not always be made explicit as an **a priori** justification for the presentation, but they underpin all but the most trivial of statistical presentations. Anyway, whether or not a writer justifies and interprets the relationships presented, readers will provide an interpretation in order to make sense of the data--or dismiss the presentation as meaningless.

Given all this, in the analysis strategy which follows we are aiming toward statistical presentations which: (a) begin with the presentation of descriptive statistics--simple displays of students' achievements; (b) move to displays, in tabular and related forms, of simple relationships defined by the model; and (c) graduate to the presentation of effect estimates.

2. Dimensions of the Task.

Someone once said 'There is nothing as practical as a good theory.' And one can easily subscribe to this view when faced with the design of an analysis plan. The theory (or theories) which gave rise to the model(s) which guide and give coherence to the study as a whole, define the structure of the analyses. One knows what variables are to be combined into constructs, what relationships are to be estimated, and what variables are to be controlled in the estimation of those relationships. The remainder is mostly technical statistical application concerned with testing the fit of the model to the data, and estimating the strength of relationships.

In this discussion the term 'theory' is taken to be synonymous with 'model'. There are two broad categories of these: *measurement models;* and *structural models.* Measurement models are theories about the way in which measured variables (responses to questionnaire or test items, for example) relate to constructs (latent or unmeasured variables such as 'alienation', or 'socioeconomic status'). Structural models are theories about the way in which constructs relate to each other. Each 'specifies the process

underlying the joint distribution of a set of observable variables'; Bielby and Hauser, 1977:141).

The problem to be faced in the design of the analysis plan for the IEA Reading Literacy study is that the measurement and structural models which gave rise to the data (whose parameters are fixed), for the most part, have not been developed explicitly. The measurement models for the reading achievement tests are best defined and these, along with some rather general structural models which attempt to capture the antecedents of individual variation in reading literacy, are all that exist on paper.

3. The General Model(s).

It is our contention that the several models which have appeared during the development of the project, though basically similar, are all too general to be meaningful as models to be estimated.

Given this argument, if there is to be a general model it ought to be comprised of constructs that relate to the reader's experience and, for the most part, these constructs will be at a lower level of abstraction. To do this, one would probably disaggregate each of the general constructs proposed into several lower level constructs, which means a lot more constructs in the model. However, the estimation of a model with a lot of variables, and the extraction of meaning from the estimates, becomes more difficult as the number of variables increases. Thus, a simple disaggregation of the constructs followed by the estimation of a single general model is probably not a particularly useful solution.

3.1. The General Model as an Organizing Principle.

While the particular model may no longer represent the current state of theoretical thinking for the study, nothing very different has been seen on paper to date. Other models proposed are all cast at the same general level of thinking. Of course, from the perspective of the international analyses, there is some convenience in having a single model that can be estimated for each participating nation and compared across nations. However, for the purposes of the U.S. analyses such general models seem to have limited utility for the definition of meaningful constructs or the estimation of effects that can be seen to have implications for social policy.

Nevertheless, we suggest that there may be value in adopting the kind of general model proposed in the various IEA documents as an organizing principle for:

(a) separate analyses which look at 'submodels' in the level of detail recommended above; and

(b) a reporting structure for these analyses.

3.1.1. As an Organizing Principle for Analyses.

The essence of this proposition is that, for reasons of meaning and statistical practicality, a collection of simpler, less abstract models should be estimated rather than a single general model. These submodels can be tied more explicitly to those questions which were formulated in advance as the justification for the study as a whole, and will provide answers which have real-world meaning to a general audience. Since the submodels are also tied together by the structure of the single general model within which they were derived, their findings contribute to a coherent picture of what it is that makes a difference to students' reading literacies. In addition, as well as providing an accounting of questions answered, such a mapping of questions and analyses has the potential to indicate where questions of interest may not have been posed.

3.1.2. As a Reporting Structure.

The overall model offers similar advantages as a general organizing principle for the reporting of the analyses. The model provides a structure for the organization of the questions and analyses to be reported, and suggests the broad divisions of the reporting task.

4. Defining Constructs; Measurement Models.

Constructs are attributes not directly experienced. Their existence is inferred from the observed covariation among observations (indicators) which we expect, on theoretical grounds, to tap an unobserved attribute. 'Narrative comprehension', for example, is a construct whose existence is inferred from the covariation among items on a test.

When constructs and their indicators are explicitly defined **a priori** and data are gathered to meet this specification, the analyses are quite straightforward. One selects an appropriate statistical technique to confirm that the theoretical relationships implied by the model are actually observed in the real world (as this is captured in the data). This is something of an ideal situation that is usually only approximated. At the other (less desirable and atheoretical) extreme is the case where one explores the data to elicit whatever latent structure there may be.

The IEA Reading Literacy study fits somewhere in between the two extremes. Since this is a collaborative enterprise, the questionnaire and test items tend to reflect the pooled theoretical/substantive knowledge and views of the National Research Coordinators (NRCs) in close to 40 nations. The NRCs are familiar, at the very least, with the study of reading, so most of the items produced tend to be pretty much on target in and of themselves. However, the items are only loosely grouped into broad categories. Such working constructs as may be captured by these item-level indicators have to be inferred from the items, after the fact. The problems here, of course, are that people's views vary a good deal about which items go together as measures of a construct, these views are not always born out empirically, constructs get unequal coverage in the data, and there is always the temptation to aggregate substantial numbers of items to form a 'construct' in order to reduce the large number of variables to a manageable size for analysis.

The tests, of course, are not a special problem. IRT procedures will elicit their latent structure and provide the rationale for the aggregation of items. However, the non-test items do pose some problems. In most cases there is no explicit measurement model to guide the analyses. And, since some kind of monster exploratory factor analysis on the total set of non-test items is not justifiable, or sensible, or feasible, the most reasonable approach seems to be one of using expert judgement to assign items to broad groupings, followed by exploratory factor analyses to allow the data to tell us what the underlying constructs are within each of the domains specified. Of course, it does not make sense to see every item as a multiple indicator of some construct; some will remain as single indicators; others will be judged to be part of a group tapping one or mare latent variables.

4.1. Item Groupings and a General Model.

We have taken the first crucial step in this process. The final internationally agreed-on set of variables have been grouped them into what we see as relatively homogeneous clusters consistent with what we know about reading and the original intent of the general IEA model. Once the data are in we will take the second step proposed above; namely, explore the latent structure of the items within each cluster. (Actually, we have made a start on this matter using the pilot-study data, and with the results one might expect; some item groupings have the latent structure expected, others do not.)

Having grouped items on a *priori* grounds into substantively meaningful clusters we took the next step as well, arranging these clusters into a general model which makes clear the nature of the relationships between clusters in a general causal model format.

5. Defining Structural Models.

The situation is basically as follows. There is a general structural model which, we have argued, is better seen as an organizing principle than a model to be estimated. There are questions that were defined in advance as those to be answered by the study. And there are any number of questions of interest to the 'reading community' as a whole, and to those associated with the IEA study itself.

If we accept the general model as an organizing principle then the task is to develop workable submodels consistent with this framework (or parts thereof) and the available data. This is the work of experts in the field--those familiar with the literature and the issues. However, there is something of a timing problem in this respect. Until the work on defining the constructs supported in these data is completed, there will be a degree of uncertainty as to what the elements of these models can be. Nevertheless, it should be possible to go ahead and define tentative operational models consistent with what might be expected from the data.

Almost certainly there will be a series of models looking at the effects of instructional practices on reading achievement, whether this varies for different aspects of reading, how instructional practices vary with different children, and whether this matters, whether questions of equity arise in this connection, and so on. Models looking at the effects of school, teacher and program characteristics would be of particular interest as these effects, along with teaching strategies, are those most amenable to change. Multilevel modelling would be part of this general emphasis as a way of looking at the effect of differences between schools in resources and in 'school climate' or ethos. And so on for the other areas of interest noted. These working models will be mapped onto the general model as a way of keeping track of the substantive issues being addressed. In this way a register of effort to be made, and effort needed, will be maintained.

D. ANALYSIS AND REPORTING

As a governing strategy we suggest that the presentation of the analyses should proceed, wherever possible, from a demonstration of relationships observed in the 'real world' to estimated effects produced by statistical analysis. Thus, each analysis has two basic components. The first of these is an examination of the **observed** relationship of each independent variable (item or construct as the situation demands), as defined by the model in question, to each of the three dependent variables--narrative comprehension, expository comprehension and document search. (In subsequent discussion the use of 'reading achievement' may be taken to mean any or all of these three variables.) The second is an examination of a parallel set of **estimated** relationships not from the 'real world', estimates of what the observed relationship would look like, **other things equal**.

1. Reporting Real World Relationships.

The rationale for this reporting strategy rests on the notion that readers are best convinced of the validity of conclusions if they can see both the data and the text as part of the same presentation. Further, they will be more accepting of unfamiliar statistical effect estimates if they can see the foundation of these estimates in the observed 'real-world' relationships among the variables in question. And, most readers will want to explore the data a little; they will want to get a basic 'feel' for the data and the study as a whole. They will be interested in seeing whether their own notions about influences on reading are born out in the data, and will want to scan the data to see what does, in fact, relate to reading achievement.

The first step in providing this statistical foundation is to suggest that each analysis, as defined by a particular submodel, should provide an overview of the basic data in the form of tables showing the relationships of items and constructs to reading achievement.

1.1. Items and Constructs.

For each model under examination, these analyses will report descriptive data on the relationship between reading achievement measures and: (a) selected items; plus(b), each appropriate construct. It is important to present item-level relationships at this early stage in the analysis as these will have concrete referents in most reader's experience, something that will be lost to varying degrees in the constructs produced. Bivariate

tabulations showing mean reading scores within categories of the item/construct, or proportions achieving some level of reading achievement are probably the form to be adopted.

The descriptive data displays are to be organized to reflect the structure of the particular model under examination. These descriptive data will form a foundation for the multivariate analyses to be developed subsequently. In the sections which follow we consider several modes of reporting these basic descriptive statistics. There are other useful forms of tabular presentations and a wide variety of graphic forms not discussed. The critical issue, of course, is to match the form of presentation with the purpose of the presentation, in the sense of the aspect of the data to be emphasized.

1.2. Reporting Reading Achievement.

The presentation strategies developed below assume that the reading literacy test items will be scaled according to an Item Response Theory model and that three scales will be involved. Our plan is that the full range of reading literacy data ranging from item level data to scale scores be used selectively in the presentation of findings in the ways outlined below.

1.2.1. Item-level presentations.

Experience suggests that a display of the specific tasks tapped by test items is of considerable interest generally. It provides a very concrete picture of what reading literacy is about. The display we intend is one in which items within any scale are ranked by difficulty with their relative difficulties defining their position on the scale. Relatively homogeneous groups of items tapping particular skills might well be indicated and proficiency cutting points determined.

1.2.2. Scale scores.

Total scores for scales provide another measure of reading literacy and one amenable to both tabular presentations and multivariate analyses. As part of the base-level reporting strategy suggested above one could consider reporting scale score means within each category of (selected) explanatory variables and constructs.

1.3. Reporting Relationships

The basic form of reporting proposed centers on conventional tabulations. We suggest as well other analogous data displays which give different perspectives on the data. These may prove effective in presenting some of the results and should be selected on this basis. The three forms of presentation proposed for the reading literacy scales suggest five forms of tabular summaries each providing a different perspective on the data.

1.3.1. Item-by-item tabulations.

This simple form of tabulation would display on an item-by-item basis the proportion of persons correct in each category of an independent variable. However, while the impact of such a concrete presentation is likely to be high, there are a large number of items to be tabulated against a large number of independent variables. Selective presentations of this kind in either tabular or graphical form could prove useful. A complete description will be provided in a companion technical volume.

1.3.2. Item-by-group mapping.

Using the kind of item-level display noted above one can map against this scale the distributions of correct responses from whatever group or groups are of interest. The result is a concretely interpretable presentation of who can do what--by ethnic group, by gender, by socioeconomic status, by nation, and the like.

1.3.3. Cross-tabulations.

Conventional cross-tabulations provide the more usual approach to presenting descriptive data on relationships. For the sake of simplicity assume that scale scores based on reading proficiency categories be used. Reading literacy defined in this way could be tabulated against variables defined as influences on achievement by the operational model under examination at the time. In each table the interpretative emphasis would be on the percentage of persons achieving each of the three proficiency levels within each category of the independent variable. Graphical representations of these same data follow in a straightforward manner.

1.3.4. Scale score means and variances.

Two forms of presentation of the observed means are suggested. The first involves 'box and whisker' plots. These graphical representations developed by Tukey (1977) allow for the presentation of the distribution of students in terms of the percentage of items correct. These data displays tell one a good deal about the form of the distribution of reading scores in each group and would be selected as the means of display for just this reason. The second form of presentation involves the display of scale score means within each of the categories of an independent variable. Such presentations lend themselves to graphic displays, and need not be confined to a single group.

2. Multivariate Analyses.

Four considerations underlie the arguments in this section. First, multivariate analyses are necessary to provide reasonably unequivocal answers to the questions which generated the IEA Reading Literacy Study. Second, for reasons already advanced, we want to have the results of multivariate analyses presented to the audience in the reports, and understood by them. Third, in order to be understood by a general audience, the results of multivariate analyses need to be presented in the statistics of common discourse--means and percentages. Fourth, while relatively complex models may underly the analyses, the findings should be reported as a series of answers to single-issue questions woven together in a storyline which reflects the basic concerns of the particular analyses, and the theoretical/substantive arguments captured in the model.

2.1. Presenting the Findings from Multivariate Analyses

The approach suggested is straightforward and computationally simple. It uses the output from standard OLS regression to produce an adjusted (dependent variable) mean for each category of the independent variable. Where the dependent variable is a dichotomy these means can be presented as proportions or percentages. Observed means for reading achievement across categories of explanatory variables will have been presented as part of the descriptive analyses. The analyses suggested here provide estimates of what these category means would be, **other things equal**. In this way we satisfy

the need to provide statistical controls on several variables simultaneously but instead of presenting the results as some form of partial regression coefficient, they are presented as adjusted means or proportions, statistics within the experience of most readers.

E. DESIGNING THE IEA REPORTS

It seems fairly clear that in designing analyses and reports for a general audience a good deal of the parsimony present in academic reporting is sacrificed in the interest of providing the audience with access to the basis for the conclusions reached.

1. Analyses

Five general principles govern the presentation of the analyses within each of the reports.

(a) First, questions to be answered by the analyses, together with hypotheses about the processes in question, will be specified in advance by those familiar with the theory and literature of reading. These issues will be woven together to provide the overall structure for the analyses in such a way that they tell a coherent story.

(b) Second, for the most part the questions/hypotheses should be single-issue in their orientation.

(c) Third, this question/hypothesis formulation must take place with the data in clear view, as the data will not allow an answer to every question or hypothesis one might like to pose. In fact, it seems that some of the questions posed in the various documents associated with the study may not be addressed by the data to be collected.

(d) Fourth, in each case the analyses should begin with the simple additive models and address more complicated propositions only if arguments are advanced **a priori** to justify a more complicated view of the world. In fact, for the bulk of the issues addressed it might be appropriate to begin with an examination of the bivariate relationship between the two variables in question and then tease out the meaning of this relationship through statistical adjustments defined by the accompanying theoretical/substantive arguments.

(e) Fifth, the findings of these analyses should be presented in text, table and graphic following the report design principles discussed above.

2. Designing Reports for a General Audience.

As argued earlier, a standard academic format for reporting is probably unsuitable in this context. Only a few members of the audience will be interested in a review of the literature and theory surrounding the issues at hand, and fewer still will be inclined to work their way through the technical details of the study. (Both of these groups will be catered for, through companion technical reports.) On the basis of experience with

reporting similar kinds of analyses to a general audience, with some success, we suggest some principles that may be of use in report design.

2.1. Models and Issues.

The specification of an overall model **a priori** is essential for the efficient design of a study, and the efficient analysis of the data that result. However, estimating and describing such a model may not be the best way to present the results of the study to a general audience. Most people are used to dealing with issues/questions/problems and the answers to issues/questions/problems. Our point is that in designing analyses for a general audience the analyses are issue- rather than model-oriented, address questions one at a time, and provide answers to these questions one at a time.

However, the model (or submodel) remains the organizing principle for the presentation of these issues and their answers. It provides the framework within which the presentation of a series of single-issue analyses is organized. This framework also provides the backbone of the storyline that is developed as data is brought to bear on the issues that generated the model in the first place.

2.2. Building Analyses.

The first point of reference in any analysis should be a straightforward display of the bivariate relationship(s) between the 'dependent' variable(s)--reading achievement--and the 'independent' variables in question. As a second step such multivariate adjustments as may be necessary (defined by theoretical considerations) should be undertaken using linear and additive models. Non-additive and/or non-linear models and/or hierarchical models could be estimated as a third step if theoretical and/or substantive considerations so dictate.

2.3. Tables, Text and Graphics.

Experience suggests that most people do not read tables, and that hardly anybody reads a lot of tables by choice. On the other hand, graphic presentations can be used with effect to highlight a particular aspect of a table. The NAEP Report Cards offer an appropriate model in this respect.

On the basis of these assumptions we suggest that the main body of the report(s) contain only a few comprehensive descriptive tabulations, that such other tables and statistics as are appropriate be provided in appendices (or in the technical report if they are only marginal to the argument), and that graphic presentations of the data be used to highlight findings of interest, anchoring the storyline to the data regularly as the story unfolds.

We suggest as well that the results of the multivariate analyses be introduced only when there are substantive arguments in the text that the observed relationships may be spurious and lead to unwarranted conclusions, or when the role of an intervening variable is the issue at hand. (The results of these analyses should appear in full in the appendices and/or the technical report.)

2.4. Catering for Methodologists and Sceptics.

Each report should be self-contained as is feasible. Some readers will want to see, in varying degrees of complexity, the empirical basis for the conclusions reported in the main text. There should be sufficient detailed supporting and supplementary material provided in appendices of each report to describe the basic 'academic' aspects of the study and the analyses to support the conclusions reached. More detailed treatments in the companion technical volumes would be referenced.

2.5. Multiple Access Routes.

Readers will have varying degrees of time to spend, and interest in, reading a report. Some will read it in total; some will dip into it for information on topics of interest; others will scan it to see what might be of interest. Each report should cater to this variety of demands on its information content.

To this end we suggest that the reports be structured in such a way that persons with varying degrees of time to spend can extract the findings in as much depth as they need. For the bulk of the audience the most detailed treatment will be the text of the report itself, though the detailed statistical basis for the report will be provided in the appendices/technical report. Marginal notes extracted directly from the text can be introduced to provide a means for reading the report in a somewhat lower level of detail; that is, by scanning down the margins to pick up the main findings. This has the advantage of exposing the reader to the text and the data presentations in the text at the same time. Where a finding is of particular interest the reader is at the spot in the report where he/she can transfer to the text itself for a more detailed exposition. Similarly, a detailed heading structure down to the level of paragraph headings allows for a little less detailed scanning of the report with the possibility of moving to marginal notes and/or the text itself as interest demands. For the busiest of readers we suggest that a collection of the marginal notes be provided at the beginning of the report as an executive summary. These should be keyed to the text by page numbers to allow the reader immediate access to the more detailed presentation. Finally, since these reports are likely to serve as something like a handbook on the issues under examination, it would be useful to provide an index for readers who wish to extract a fact or finding quickly without scanning through the whole report.

F. THE SUBSTANCE AND STRUCTURE OF THE U.S. NATIONAL REPORT.

Within the framework developed (Figure 2), and with the foregoing considerations in mind, we have developed the broad outline of the story to be told in the U.S. national report. The backbone of the argument is a series of 'commonsense' questions/explanations/issues which we might expect people to raise if we asked them why it is that U.S. students vary in their reading achievements. These 'commonsense' issues are only a subset of all commonsense questions that one might ask in this context; however, they are that subset which can be addressed with the IEA data as mapped onto the model shown in Figure 2. In other words, they are questions which can be addressed with the ICC-defined data that will become available next year. By way of illustration, in the outline of the third section of the report, questionnaire items pertinent to each issue are indicated by their questionnaire number. Obviously, this is not an exhaustive list of

questions that one might ask within the boundaries of the ICC-defined data. An outline of the proposed 11 sections of the national report follows.

1. Reading Literacy in the United States

Describe U.S. performance on tests; by males, females, persons (using NAEP model?); item and scale displays in some detail; interpretations of the data in terms of reading theory?.

2. The U.S. in International Perspective

Describe U.S. performance *vis a vis* other nations; concentrate on nations which are U.S. competitors in the world markets, and on nations with strong reading programs, such as New Zealand; pick the highlights out of the international analyses, and/or conduct some of our own to address questions of particular interest within the U.S.; but not a full-blown exposition that duplicates what the ICC will produce.

3. Developing Competence in Reading Comprehension

Develop the 'big' issues in common-sense terms anchored in these data displays. Tell a story to an interested but substantively and technically uninformed audience. Tell the story as a series of questions that the educated lay-person might pose. The purpose of the section is to set up the case which justifies and structures the analyses to follow. Splice in a review of the literature and published statistics here around each of the issues.

Some of the simpler and more obvious questions are outlined below. The questionnaire items which pertain directly to the issue are noted by reference to the map provided separately.

a. Why do some students read and comprehend well, and others do not? What do we know and/or suspect about the processes at work?
A general introduction to the literature, theory and substantive knowledge about reading to provide a foundation for the questions and analyses which follow.

b. Are there subpopulation groups who appear to be disadvantaged?

the poor?

the least educated?

ethnic minority groups?

families in rural areas?

inner-city families?

c. Is this something to do with what goes on inside homes between parents and children?

Do poor readers watch too much TV?

Is it because some families can't afford to provide the resources needed?

Do family/work responsibilities interfere with learning?

Is it because there are subcultu al differences involved--in the emphasis placed on reading, and the degree of parental support provided, or the habits of parents which provide models for the children, and/or simply exposure to amounts and variety of print material?

If another language is used, does this matter? How much 'other language' can students deal with before it affects their reading? Are bilingual students better off than monolingual students in either language?

Are there gender differences in these effects?

d. Maybe it's the schools. Are some schools better than others?

Are small schools better? or private schools? or urban schools?

And are any differences we might find really due to the schools?

or are they due to the fact that there is social and economic segregation in the community such that there are 'poor' schools,

and 'low education' schools,

and 'minority' schools, and so on.

We would have to take this into account. If we do that and we still find differences between schools then,

are some schools better than others because they have more physical resources;

or better community resources?

or better teacher resources in the sense of better trained teachers;

or maybe they simply spend more time overall teaching the students?

Or they have special programs for reading?

Maybe it's to do with the kind of leadership provided by the principal;

perhaps it's to do with parent involvement in the school?

and then there is the whole question of 'school/classroom climate', the effects of student-body characteristics in the aggregate

e. Maybe it has a lot to do with the kind of teachers that teach reading;

are there big differences in reading comprehension between classrooms?

does the education and experience of the teacher matter?

is teacher gender *vis a vis* student gender important?

what about teacher ethnicity *vis a vis* student ethnicity?

are the teacher's own reading interests influential in this context?

do the teacher's views on the aims of reading instruction make a difference?

do teacher/parent meetings help? and the teacher's attitudes to reading instruction?

are the teacher's expectations about the student's own capabilities important?

f. Maybe we should have a good look at what teachers do;

their general instructional strategies;

the specific reading/comprehension strategies they use;

how much time they spend teaching reading;

how much they have in the way of reading resources in the classroom;

the extent to which they have children in need of help;

and the extent to which these students get this help.

does the amount of homework they set matter?

does instructional grouping have an effect?

hat about silent reading and written work--do they matter?

is it better if the lessons are teacher-directed or student-directed?

is it all a matter of the amount of time they devote to the three kinds of text (narrative, expository, document)?

where do the assessment practices of teachers fit into the picture? is it important how often this is done? or by what method?

are the teacher's own views on the aims of reading instruction important for student success?

q. But students have a mind of their own, as well.

Do the kinds of material they read at school and outside school make a difference to their reading comprehension in each of the three areas?

is their opinion of their own reading capabilities important?

what about their educational expectations generally?

To what extent are these student characteristics a function of what teachers, and schools and families do?

4. Is the Locus of Disadvantage in Families?.

The measures available allow a focus on three subpopulation groups; those defined by *social* status (father's and mother's education); *economic* status (family wealth measures); and *ethnicity*.

Display the relationship of each to the reading literacy measures in tabular form and offer preliminary interpretations of the tables; are there between-group differences? how big are they? are they worth worrying about? do we know what causes them?

Now take up the question of whether these differences come about because of between-group differences in family environments; that is, control for the following as intervening variables--family wealth, family language, print exposure, parent interest/involvement, student's responsibilities (job etc.)
For example, if we are able to demonstrate ethnic group differences in reading achievement, we might ask ourselves whether these differences come about because there are parallel ethnic group differences in parental interest and involvement in the child's reading activities at home--expressed a little differently, something like 'subcultural' differences in the value placed on these capabilities? subcultural differences in knowledge of what appropriate parental behaviors may be? and so on. One way of looking at this question is to: first, measure the relationship between ethnicity and reading achievement; and second, measure this same relationship with statistical controls on parental interest/involvement. (This can be done in tabular presentations or via multivariate statistical techniques based on the general linear model.) If the second (adjusted) relationship between ethnicity and achievement is smaller than the first, and given that our model is correct in postulating that parental interest is an intervening variable, then the subcultural differences argument is supported. The actual analyses are likely to be more elaborate than this as there are several reading achievement measures and 10 or so measures of parental interest and involvement. However, the general principle still holds.

Multi-way tabulations might be used in such analyses but for presentation purposes they would probably get a bit unwieldy, and certainly would take up a lot of space. It should be possible to start talking about observed and adjusted means at the very beginning, with the view to providing analyses whose interpretations rely on changes in observed and adjusted slopes, as notewd above. What is probably best is to weave together a mixture of tables and graphics of this sort.

As a next step we might show in a similar way how the several family environment measures relate to reading achievement--tables and graphs--and start to talk about the relative importance of these family variables in and of themselves, and as the explanation for the observed disadvantage suffered by some subpopulation groups (assuming we find some).

The point of all this is to build up a story using tables and graphs to display the data. We would not display everything, but rather select the pertinent pieces of the data to support the story being told. The story, of course, is about demonstrating the extent of socioeconomic and ethnic group differences in reading achievement, and then providing evidence of the socioeconomic and ethnic differences in family environments that are responsible, at least in part.

An example of this kind of development of tabular, multivariate statistics and graphic presentations around a particular issue is provided as attachment #2. The issue here is participation in education rather than reading literacy but the arguments are much the same. While the presentation is a fraction primitive in this instance the general idea is that the analyses, as presented, address single issues. Each issue is anchored in tabular presentations of the data--a form readily understood. Aspects of these data may be highlighted in graphics. The multivariate statistics are developed because they are *necessary* to provide a reasonably unequivocal answer to the question under consideration, and are presented in the form of statistics that are familiar to most readers.

4.1. Policy Implications.

At this point in each section we would want to give some thought to the policy implications of what we have found. Some of the policy implications we might arrive at will be suggestions consistent with the findings but not directly derivable from them. For example, if we should find that children from 'poor' families are less competent in reading apparently because they have few reading resources in their homes, we might argue that programs of direct compensation might work--means-tested payments to families, vouchers to buy books, special library programs designed to get books into these homes, and such. The data will not allow us to say definitively that these will work to reduce reading disadvantage, but it's a reasonable bet. (Family intervention programs are, of course, administratively and politically cumbersome. The easiest and traditional route by far is to use the schools.)

So, it would be useful to consider whether we could use the IEA data itself to bring more direct evidence to bear on this matter. It might be possible to address these questions by analyzing other data sets as well. That is, on the basis of these findings we could think about the design of further analyses within the IEA data, and/or within other data sets such as HS&B or NELS:88, analyses to provide some kind of handle on the ways in which schools might compensate for family disadvantage. Are some kinds of schools and/or teachers better able to compensate for these disadvantages than others? Do students from 'poor' families do better in some schools and/or some school situations than others? Are there particular instructional practices better suited to such students than others? It would be sensible to set the various recommendations arising from this process within the context of reading theory and substantive knowledge.

5. Schools and Reading Achievement

Adopt a similar approach. Start off with a display of the data to establish whether there is something to talk about; tabulations of school size, public/private, and urban/rural schools against the three measures of reading achievement. Then start to wonder, as before, about the reasons for these differences. Is it simply differences in the student body?--adjust for student characteristics. Is it more than this? If so, think about resource differences between schools, and adjust for these, either in n-way tabulations or with multivariate adjustments to produce adjusted means and such.

We could do the same for leadership within schools. Ten or more of the measures we have classified under 'school environment' pertain to the activities of the school principal as reported by both the principal and

the teacher. So we could take a look at the way in which variation in school leadership influences reading achievement through what amounts to a measure of school climate on some kind of control--autonomy dimension. Again, it would be a matter of selecting out of these analyses what matters, and weaving the story around these items (and a few items that we thought mattered, and appear not to) with selective data displays to reinforce the words.

6. Teachers in Classrooms

The pertinent data have to do mainly with teacher education/training, ethnicity, the characteristics of the class, the reading resources available within the classroom, and the amount of time spent on language teaching and/or the teaching of reading. A set of analyses looking at the relative effects of teachers as resources together with reading material resources within classrooms could be enlightening. Questions to do with teacher attitudes towards reading and towards the students themselves could be raised in this context--and answered, at least in part. So too could some longstanding questions regarding the effects of congruence in student/teacher ethnicity and gender.

In this section we might want to address as well the question of the effects of school and classroom 'climate', as this is measured by aggregate student characteristics--for example, the effect of being in a classroom where most of the other children are wealthy (or poor), or are from minority ethnic groups (or are from the majority group, in the case of minority students), and so on. In this section we would also capitalize on the fact that the sampling plan provides for data on two classrooms within each elementary school. The analyuses in question would focus on the question of between-school and between-classroom differences in reading achievement.

7. Instruction in Reading

The data offer a good opportunity to develop an elaborate and detailed treatment of the effects of teachers' instructional practices on students' achievements. There are a number of measures of teachers' general instructional practices, the content of their instruction, and the kinds of strategies they use in the teaching of reading. There is opportunity here to pick up on the matters raised under 'Policy Implications' above and look at questions about different practices and strategies suit different kinds of students--something similar to the aptitude-by-treatment interaction model.
Similarly, one could take a detailed look at the way in which teacher assessment practices fit into the picture. There are 15 or so measures of such practices in the data.

8. Student Behaviors and Attitudes

The most extensive set of measures in this category concern students' reading activities--the amount of time they spend reading various forms of print relating to school and homework, and for leisure. This might well be considered in chapter 4 in connection with the investigation of family influences.

9. Differences between Populations

By looking at differences in the size and patterns of effects across the two populations it may be possible to get a handle on age differences in these processes; to examine, for instance, whether factors important in the reading achievement of 10-year-olds in elementary school remain important for the reading achievements of 14-year-olds. There will be some difficulties in establishing the comparability of the measures across the two populations; however, some qualified analyses seem possible and are likely to offer findings pertinent to the overall picture of reading achievement that is being developed.

10. Differences between the U.S. and other Nations

Depending on what the international analyses look like, it might be a good idea here to bring in some international comparisons of the social processes that affect reading achievement. Selected comparisons of the kind of analyses outlined above.

11. What Makes a Difference?

In each of the chapters/sections/whatever above, interpretations, conclusions, and policy implications specific to the concerns of the section would have been developed. The point of this final chapter is to pick these up and stitch them together to tell everyone what works as far as reading comprehension is concerned, and what we ought to be doing as a result to raise the general level of reading comprehension, and reduce the disadvantage suffered by some subpopulation groups. Thus, the chapter would summarize the whole set of analyses in something like an extended executive summary and, in so doing, tie together all the findings and recommendations in one coherent and integrated statement on reading literacy in the U.S.